

#6



OIEP

RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/816,669A

TIME: 13:05:26

Input Set : A:\EP.txt

Output Set: N:\CRF3\02102002\I816669A.raw

ENTERED

OF

3 <110> APPLICANT: GARABEDIAN, Michael
 4 TANEJA, Samir
 5 HITTELMAN, Adam
 6 MARKUS, Steven
 8 <120> TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS
 9 TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL COREGULATORY
 10 PROTEINS AS TARGETS FOR ANDROGEN RECEPTOR-DEPENDENT DISEASES
 12 <130> FILE REFERENCE: GARABEDIAN=1.1A
 14 <140> CURRENT APPLICATION NUMBER: 09/816,669A
 15 <141> CURRENT FILING DATE: 2001-03-26
 17 <150> PRIOR APPLICATION NUMBER: 60/225,618
 18 <151> PRIOR FILING DATE: 2000-08-15
 20 <150> PRIOR APPLICATION NUMBER: 60/191,768
 21 <151> PRIOR FILING DATE: 2000-03-24
 23 <160> NUMBER OF SEQ ID NOS: 20
 25 <170> SOFTWARE: PatentIn version 3.1
 27 <210> SEQ ID NO: 1
 28 <211> LENGTH: 474
 29 <212> TYPE: DNA
 30 <213> ORGANISM: human
 32 <400> SEQUENCE: 1
 33 atggcgacgc cccctaagcg gcgggcggtg gaggccacgg gggagaaagt gctgcgctac 60
 35 gagaccttca tcagtgcagt gctgcagcgg gacttgcgaa aggtgctgga ccatcgagac 120
 37 aaggtatatg agcagctggc caaatacctt caactgagaa atgtcattga gcgactccag 180
 39 gaagctaagc actcggagtt atatatgcag gtggatttgg gctgtaactt cttcgttgac 240
 41 acagtgggtcc cagatacttc acgcattctat gtggccctgg gatatggttt tttcctggag 300
 43 ttgacactgg cagaagctct caagttcatt gatcgtgaaga gctctctcct cacagagctc 360
 45 agcaacagcc tcaccaagga ctccatgaat atcaaagccc atatccacat gttgctagag 420
 47 gggcttagag aactacaagg cctgcagaat ttccagaga agcctcacca ttga 474
 50 <210> SEQ ID NO: 2
 51 <211> LENGTH: 157
 52 <212> TYPE: PRT
 53 <213> ORGANISM: human
 55 <400> SEQUENCE: 2
 57 Met Ala Thr Pro Pro Lys Arg Arg Ala Val Glu Ala Thr Gly Glu Lys
 58 1 5 10 15
 61 Val Leu Arg Tyr Glu Thr Phe Ile Ser Asp Val Leu Gln Arg Asp Leu
 62 20 25 30
 65 Arg Lys Val Leu Asp His Arg Asp Lys Val Tyr Glu Gln Leu Ala Lys
 66 35 40 45
 69 Tyr Leu Gln Leu Arg Asn Val Ile Glu Arg Leu Gln Glu Ala Lys His
 70 50 55 60
 73 Ser Glu Leu Tyr Met Gln Val Asp Leu Gly Cys Asn Phe Phe Val Asp

RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/816,669A

TIME: 13:05:26

Input Set : A:\EP.txt

Output Set: N:\CRF3\02102002\I816669A.raw

```

74 65          70          75          80
77 Thr Val Val Pro Asp Thr Ser Arg Ile Tyr Val Ala Leu Gly Tyr Gly
78          85          90          95
81 Phe Phe Leu Glu Leu Thr Leu Ala Glu Ala Leu Lys Phe Ile Asp Arg
82          100          105          110
85 Lys Ser Ser Leu Leu Thr Glu Leu Ser Asn Ser Leu Thr Lys Asp Ser
86          115          120          125
89 Met Asn Ile Lys Ala His Ile His Met Leu Leu Glu Gly Leu Arg Glu
90          130          135          140
93 Leu Gln Gly Leu Gln Asn Phe Pro Glu Lys Pro His His
94 145          150          155
97 <210> SEQ ID NO: 3
98 <211> LENGTH: 1097
99 <212> TYPE: DNA
100 <213> ORGANISM: human
102 <400> SEQUENCE: 3
103 aaatgcacaa cccggacgga agtgcctctc cgacagcaga tccaggctcg gagctccaga 60
105 cgctgggaca ggcgcgccgc agaccacccc cgccgcgcgc gggacacgac gccccccgca 120
107 ggacacgccc atcagcccgg aaaccctga gctgcttctc ccggaggccg atgcccaccc 180
109 gggagccccc aaagactcgc ggctcccggg ggcacctgca tactcaccgc cctgggcctg 240
111 ggcgcccgct gcagggactg gcgcgccgag gcctcaaaac cagcgccccc cgccctccgt 300
113 gccagcccca gccgggaccc cacaaggcaa agaccaagaa gattgtgttt gaggatgagt 360
115 tgctctccca ggccctcctg ggcgcgaaga agcctattgg agccatccct aaggggcata 420
117 agcctaggcc ccacccagtg cccgactatg agcttaagta cccgccagtg agcagtgaga 480
119 gggaaacggag ccgctatgtc gcagtgttcc aggaccagta cggagagttc ttggagctcc 540
121 agcacgaggt ggggtgtgca caggcaaagc tcaggcagct ggaggccctg ctgagctccc 600
123 tgccccacc ccaaagccag aaggaggccc aagttagcag ccgggtttgg agggagtttg 660
125 agatgaagcg aatggatcct ggcttcctgg acaagcaggc tcgctgccac tacctgaagg 720
127 gtaaaactgag gcattctcaag actcagatcc agaaattcga tgaccaagga gacagcgagg 780
129 gctccgtgta cttctaagtg cccctgcaga tgggcagagg gatgcatggg gatgcaggtc 840
131 ccttgcatth cttggtatct ctcagctttt cctcttgcat cccccctac caggggtcgc 900
133 tttctcctgg attgcaaatt cctcttcagt ttggactcag ctctgacagc cctcctcca 960
135 ggaaggcctt ccaggacttc ctctctgagg tcctctagct ctgaccctac agggactcca 1020
137 gatctcaacc tgttccctgg aagtagggcc tgctctccat cccagtgaat taaacatgta 1080
139 ttagacacct aaaaaaa
142 <210> SEQ ID NO: 4
143 <211> LENGTH: 264
144 <212> TYPE: PRT
145 <213> ORGANISM: Human
147 <400> SEQUENCE: 4
149 Met His Asn Pro Asp Gly Ser Ala Ser Pro Thr Ala Asp Pro Gly Ser
150 1          5          10          15
153 Glu Leu Gln Thr Leu Gly Gln Ala Ala Arg Arg Pro Pro Pro Pro Arg
154          20          25          30
157 Ala Gly His Asp Ala Pro Arg Arg Thr Arg Pro Ser Ala Arg Lys Pro
158          35          40          45
161 Leu Ser Cys Phe Ser Arg Arg Pro Met Pro Thr Arg Glu Pro Pro Lys
162          50          55          60
165 Thr Arg Gly Ser Arg Gly His Leu His Thr His Pro Pro Gly Pro Gly

```

RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/816,669A

TIME: 13:05:26

Input Set: A:\EP.txt

Output Set: N:\CRF3\02102002\I816669A.raw

```

166 65          70          75          80
169 Pro Pro Leu Gln Gly Leu Ala Pro Arg Gly Leu Lys Thr Ser Ala Pro
170          85          90          95
173 Arg Pro Pro Cys Gln Pro Gln Pro Gly Pro His Lys Ala Lys Thr Lys
174          100          105          110
177 Lys Ile Val Phe Glu Asp Glu Leu Leu Ser Gln Ala Leu Leu Gly Ala
178          115          120          125
181 Lys Lys Pro Ile Gly Ala Ile Pro Lys Gly His Lys Pro Arg Pro His
182          130          135          140
185 Pro Val Pro Asp Tyr Glu Leu Lys Tyr Pro Pro Val Ser Ser Glu Arg
186 145          150          155          160
189 Glu Arg Ser Arg Tyr Val Ala Val Phe Gln Asp Gln Tyr Gly Glu Phe
190          165          170          175
193 Leu Glu Leu Gln His Glu Val Gly Cys Ala Gln Ala Lys Leu Arg Gln
194          180          185          190
197 Leu Glu Ala Leu Leu Ser Ser Leu Pro Pro Pro Gln Ser Gln Lys Glu
198          195          200          205
201 Ala Gln Val Ala Ala Arg Val Trp Arg Glu Phe Glu Met Lys Arg Met
202          210          215          220
205 Asp Pro Gly Phe Leu Asp Lys Gln Ala Arg Cys His Tyr Leu Lys Gly
206 225          230          235          240
209 Lys Leu Arg His Leu Lys Thr Gln Ile Gln Lys Phe Asp Asp Gln Gly
210          245          250          255
213 Asp Ser Glu Gly Ser Val Tyr Phe
214          260

```

217 <210> SEQ ID NO: 5

218 <211> LENGTH: 517

219 <212> TYPE: DNA

220 <213> ORGANISM: Human

222 <220> FEATURE:

223 <221> NAME/KEY: misc_feature

224 <222> LOCATION: (65)..(65)

225 <223> OTHER INFORMATION: n at position is unknown.

228 <400> SEQUENCE: 5

WOK

```

229 gaacggcacg agggcgcgcc acgcgcgggg agcgggcgcg ggagcgcgcg cggcggggcc 60
231 cgcanccgag ggagccgagc gcccgmacgc gcccgagcgg acasacgcca gagccgcgcc 120
233 ccggggccgag cgcagcgcgc cgcccgssyg ggccgccagg ggcgcgcgcg gcggagcgcg 180
235 gggcgcgmgg aaagggggccc ggcgagacc aagggcaggc gcggcccgca agggcgccgg 240
237 ggaagggcgcc cggcaaggag gcggacaagc ggagcaggcc aacgagacgc gcgcacccac 300
239 acacgagcgcg gagccgccac aacaccacac ccggcccaag gagaacagca cgccaacgcg 360
241 ccagycacgg cgggcacggg aggcggggcca cacacagcgg ccccgccaag gcacggcgca 420
243 cggcacaagg gcaccacgcc agacaagcga ggaggcagca cgccgagacc ggccggaggg 480
245 ccgcgaccgc cggagaaaag gaacagagag cccccc 517

```

248 <210> SEQ ID NO: 6

249 <211> LENGTH: 189

250 <212> TYPE: PRT

251 <213> ORGANISM: Human

253 <400> SEQUENCE: 6

255 Glu Phe Gly Thr Arg Ala Arg Phe Thr Arg Gly Lys Ser Ala Leu Leu

RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/816,669A

TIME: 13:05:26

Input Set : A:\EP.txt

Output Set: N:\CRF3\02102002\I816669A.raw

```

256 1          5          10          15
259 Glu Arg Ala Leu Ala Arg Pro Arg Thr Glu Val Ser Leu Ser Ala Phe
260          20          25          30
263 Ala Leu Leu Ser Pro Ser Trp Tyr Ser Thr Ala Arg Ala Val Phe Ser
264          35          40          45
267 Val Ala Glu Leu Gln Ser Arg Leu Ala Ala Leu Gly Arg Gln Val Gly
268          50          55          60
271 Ala Arg Val Leu Asp Ala Leu Val Ala Arg Glu Lys Gly Ala Arg Arg
272 65          70          75          80
275 Glu Thr Lys Val Leu Gly Ala Leu Leu Phe Val Lys Gly Ala Val Trp
276          85          90          95
279 Lys Ala Leu Phe Gly Lys Glu Ala Asp Lys Leu Glu Gln Ala Asn Asp
280          100         105         110
283 Asp Ala Arg Thr Phe Tyr Ile Ile Glu Arg Glu Pro Leu Ile Asn Thr
284          115         120         125
287 Tyr Ile Ser Val Pro Lys Glu Asn Ser Thr Leu Asn Cys Ala Ser Phe
288          130         135         140
291 Thr Ala Gly Ile Val Glu Ala Val Leu Thr His Ser Gly Phe Pro Ala
292 145         150         155         160
295 Lys Val Thr Ala His Trp His Lys Gly Thr Thr Leu Met Ile Lys Phe
296          165         170         175
299 Glu Glu Ala Val Ile Ala Arg Asp Arg Leu Glu Gly Arg
300          180         185

```

303 <210> SEQ ID NO: 7

304 <211> LENGTH: 126

305 <212> TYPE: DNA

306 <213> ORGANISM: Human

308 <400> SEQUENCE: 7

309 gaattcggca cgaggctcaa gccctacgtg agctacctcg cccctgagag cgaggagacg 60

311 cccctgacgg ccgcgcagct cttcagcaag ccgttgccgc cttgccatcg aaaaggactt 120

313 caagga 126

316 <210> SEQ ID NO: 8

317 <211> LENGTH: 42

318 <212> TYPE: PRT

319 <213> ORGANISM: Human

321 <400> SEQUENCE: 8

323 Glu Phe Gly Thr Arg Leu Lys Pro Tyr Val Ser Tyr Leu Ala Pro Glu

324 1 5 10 15

327 Ser Glu Glu Thr Pro Leu Thr Ala Ala Gln Leu Phe Ser Lys Pro Leu

328 20 25 30

331 Ala Pro Cys His Arg Lys Gly Leu Gln Gly

332 35 40

335 <210> SEQ ID NO: 9

336 <211> LENGTH: 678

337 <212> TYPE: DNA

338 <213> ORGANISM: Human

340 <220> FEATURE:

341 <221> NAME/KEY: misc_feature

342 <222> LOCATION: (651)..(651)

RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/816,669A

TIME: 13:05:26

Input Set : A:\EP.txt

Output Set: N:\CRF3\02102002\I816669A.raw

343 <223> OTHER INFORMATION: n at position is unknown.

346 <220> FEATURE:

347 <221> NAME/KEY: misc_feature

348 <222> LOCATION: (657)..(657)

349 <223> OTHER INFORMATION: n at position is unknown.

352 <400> SEQUENCE: 9

353 gaattcggca cgaggattca ttgccccac aatcctaggc ctaccgcgcg cagtactgat 60
 355 cattctatatt cccctctat tgatccccac ctccaaatat ctcatcaaca accgactaat 120
 357 caccacccaa caatgactaa tcaaactaac ctcaaaacaa atgataacca tacacaacac 180
 359 taaaggacga acctgatctc ttatactagt atccttaato atttttattg ccacaactaa 240
 361 cctcctcgga ctccctgcctc actcatttac accaaccacc caactatcta taaacctagc 300
 363 catggccatc cccttatgag cgggcgcagt gattataggc ttctgctota agattaaaaa 360
 365 tgccttagcc cacttcctac cacaaggcac acctacaccc cttatcccca tactagttat 420
 367 tatcgaaacc atcagcctac tcattcaacc aatagccctg gccgtacgcc taaccgctaa 480
 369 cattaactga ggccacctac tcatgcacct aattggaagc gccaccctag caatatcaac 540
 371 cattaacctt cctctacact tatcatcttc acaattctaa ttctactgac tatcctagaa 600
 OK-> 373 atcgctgtcg ccttaattcca agcctacgtt ttcacacttc tagtaagcct ntactgnacg 660
 375 acaacacata aaaaaaaaaa 678

378 <210> SEQ ID NO: 10

379 <211> LENGTH: 60

380 <212> TYPE: PRT

381 <213> ORGANISM: Human

383 <400> SEQUENCE: 10

385 Glu Phe Gly Thr Arg Ile His Cys Pro His Asn Pro Arg Pro Thr Arg

386 1 5 10 15

389 Arg Ser Thr Asp His Ser Ile Ser Pro Ser Ile Asp Pro His Leu Gln

390 20 25 30

393 Ile Ser His Gln Gln Pro Thr Asn His His Pro Thr Met Thr Asn Gln

394 35 40 45

397 Thr Asn Leu Lys Thr Asn Asp Asn His Thr Gln His

398 50 55 60

401 <210> SEQ ID NO: 11

402 <211> LENGTH: 1918

403 <212> TYPE: DNA

404 <213> ORGANISM: Human

406 <400> SEQUENCE: 11

407 gaattccaat gtggttaaagt ctctgctcaa acatcacaaac ttgcaaggca ttggagagtt 60
 409 catactggag aaaaacctta caagtgtaat gactgtggca gagccttttag tgatcgttca 120
 411 agcctaactt ttcatcaggc aatacatact ggagagaaac cttacaaatg tcatgaatgc 180
 413 ggcaaggttt ttaggcacaa ttcatacctt gcaactcatc ggcgaattca tactggagag 240
 415 aaaccttaca agtgtaatga gtgtgggaaa gccttttagta tgcattcaaa cctaactacc 300
 417 cataaggtca tccatactgg agagaagcct tacaaatgta atcaatgtgg caaggtcttc 360
 419 actcagaact cacaccttgc aaatcatcaa aggactcaca ccggagagaa accttaccga 420
 421 tgcaatgagt gtgggaaagc ctctcagtgtt cgttcaagcc taaccaccca tcaggcaatc 480
 423 catactggga aaaaacctta caaatgtaat gaatgtggca aggtctttac tcaaaatgct 540
 425 cacctggcaa atcaccgaag aattcatact ggggagaaac cttacaggtg tacagagtgt 600
 427 gggaaagcct ttagggtaag atcaagtcta actaccata tggcaatcca cactggagaa 660
 429 aagcgttaca aatgtaatga gtgtggcaag gtcttcaggc agagttcaaa tcttgcaagt 720
 431 catcacagaa tgcataccgg agagaaacct tacaaatgag tgtggtgagg tcattaggta 780

VERIFICATION SUMMARY

DATE: 02/10/2002

PATENT APPLICATION: US/09/816,669A

TIME: 13:05:27

Input Set : A:\EP.txt

Output Set: N:\CRF3\02102002\I816669A.raw

L:231 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5

L:373 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9